Hypervisor-based Prototyping of Disaggregated Memory and Benefits for VM Consolidation

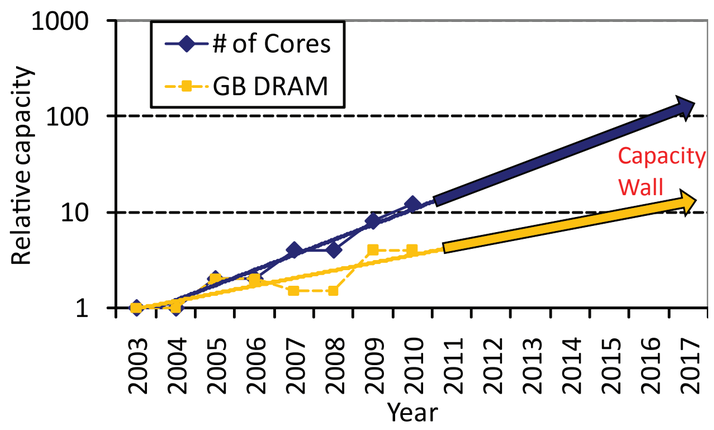
Kevin Lim[1], Jichuan Chang[2], Jose Renato Santos[2], Yoshio Turner[2], Trevor Mudge[1], Parthasarathy Ranganathan[2], Steven Reinhardt[1,3], Thomas Wenisch[1]

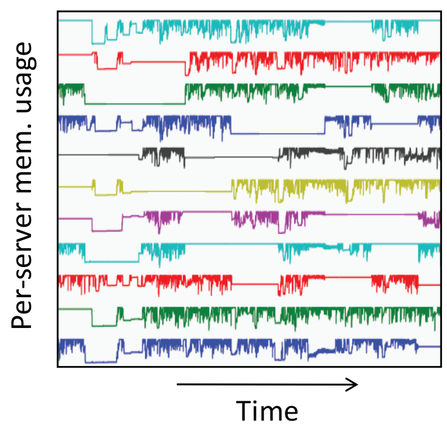
[1]University of Michigan, Ann Arbor [2]Hewlett-Packard Labs [3]Advanced Micro Devices

1 Introduction

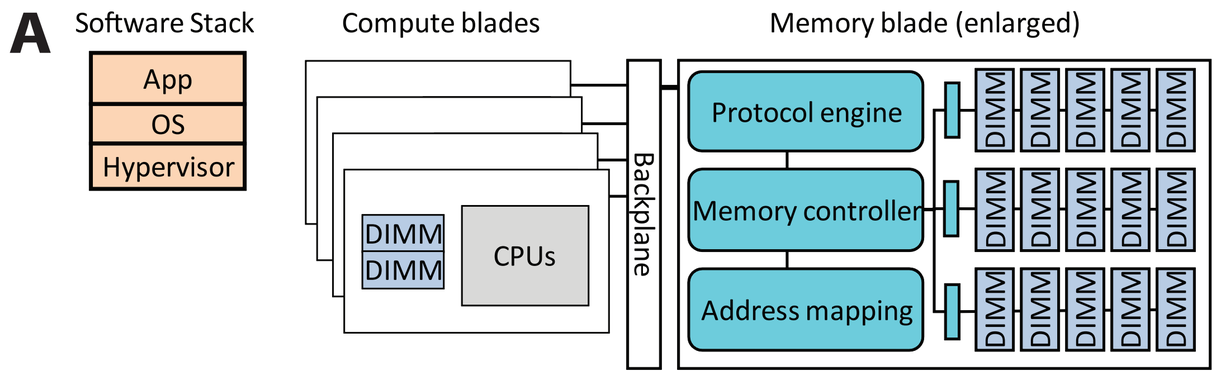
Problem: Impending memory capacity wall for commodity systems!

Opportunity: Optimizing for ensemble-level memory usage.





Proposed Solution: Disaggregated memory to provide memory capacity expansion and dynamic capacity sharing.[1]



2 Prototype

Goal: Explore system-level implications of disaggregated memory.
What hypervisor changes are needed?
Are new system-level optimizations enabled?

Current Research: Implementing a Xen hypervisor-based prototype of disaggregated memory (Figure B).
The majority of the code changes are in:
(1) memory management
(2) user-level tools for setting parameters

Functionality: Detects accesses to remote memory, on remote access swaps remote page with local page.

Configuration: 2x Quad-core AMD Opteron 2354, 32 GB DRAM. Large memory workloads: SPEC CPU (perlbench, gcc, bwaves, mcf, zeusmp), PARSEC (blackscholes), quicksort, TPC-H.

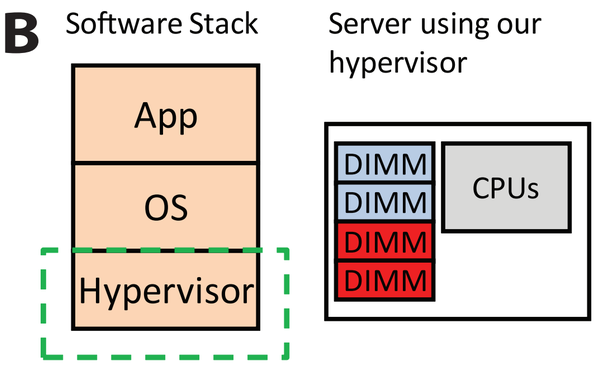


FIgure A. Disaggregated memory architecture, with memory blade shared by multiple compute blades.

Figure B. Architecture of implemented prototype. Modifications are made to hypervisor; part of local memory emulates remote memory.

3 Results

Virtual machines (VM) are setup with 2 or 8 GB of local and remote memory. Percentage and latency of remote memory are varied across runs. Results with the hypervisor-based prototype, Figure C and D, show disaggregated memory provides remote capacity with minimal slowdown.

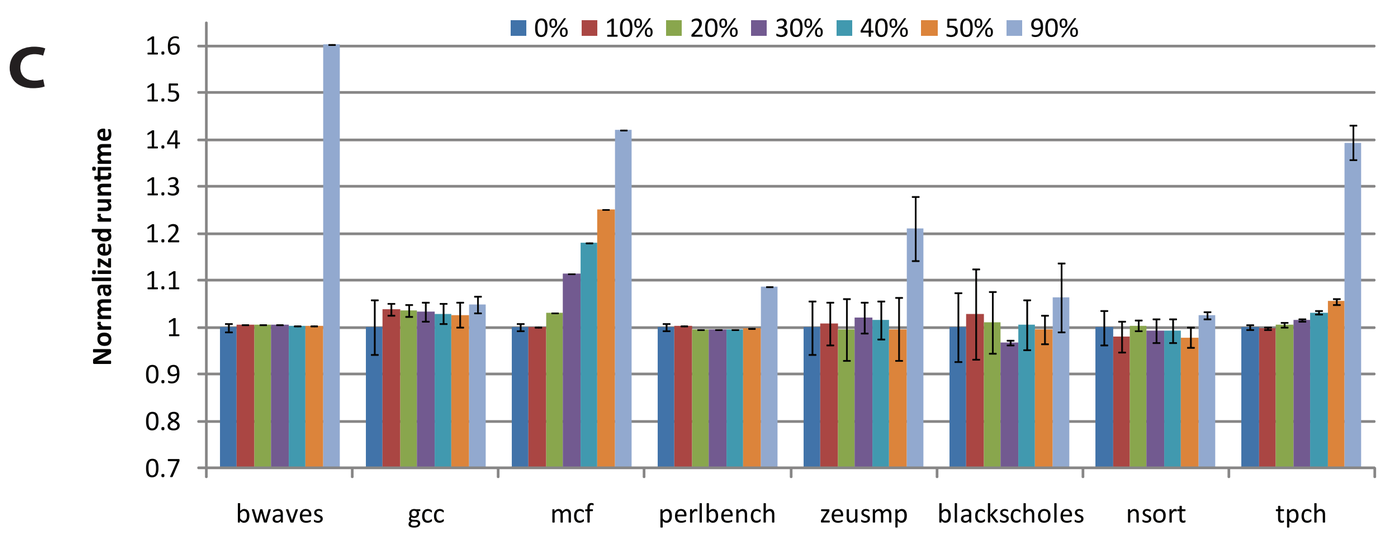


Figure C. Performance of workloads as percentage of remote memory is varied. SPEC workloads have 2 GB of memory, all others have 8 GB.

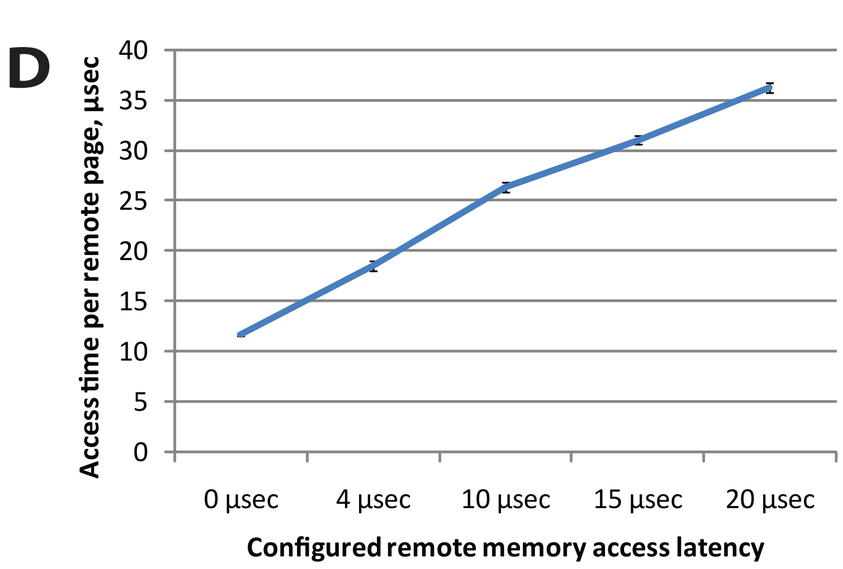


Figure D. Microbenchmark results showing measured remote memory latency versus configured latency.

4 Virtual Machine Consolidation

Disaggregated memory offers independent memory and CPU scaling, removing the memory bottleneck for VM consolidation. Other benefits are possible through techniques to increase effective memory capacity, such as content-based page sharing (CBPS).

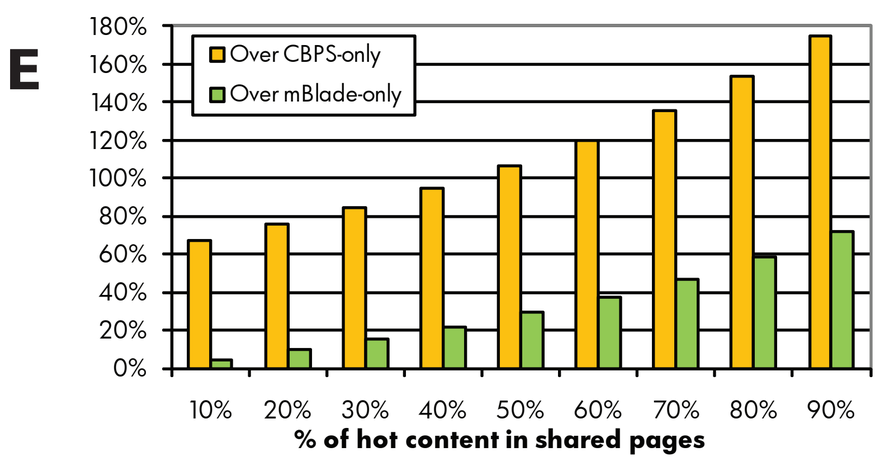


Figure E. Potential VM consolidation savings enabled by memory disaggregation with CBPS at the memory blade level.

5 Conclusions

Memory disaggregation addresses the capacity wall while providing greater efficiency and OS-transparency. Our prototype shows the feasibility of implementing disaggregated memory in a widely-used open source hypervisor. Importantly, it enables in-depth research into large-scale workloads and content-based page sharing.

6 Future Work

We plan to use the prototype to model multiple systems sharing a memory blade. We will also explore coordinating page migration with VM scheduling for greater consolidation. One use case is for virtual desktops, where non-shared pages can be evicted when a VM is descheduled (user think time), and the evicted pages are prefeteched when idle VM is rescheduled.

[1] K. Lim, et al., "Disaggregated memory for expansion and sharing in blade servers," Proceedings of the 36th Annual International Symposium on Computer Architecture, June 2009